

## Elements of fractal geometry in the $^1\text{H}$ NMR spectrum of a copolymer intercalation-complex: Identification of the underlying Cantor set

Article (Accepted Version)

Shaw, John S, Vaiyapuri, Rajendran, Parker, Matthew P, Murray, Claire A, Lim, Kate J C, Pan, Cong, Knappert, Marcus, Cardin, Christine J, Greenland, Barnaby W, Grau-Crespo, Ricardo and Colquhoun, Howard M (2018) Elements of fractal geometry in the  $^1\text{H}$  NMR spectrum of a copolymer intercalation-complex: Identification of the underlying Cantor set. *Chemical Science*, 9 (17). pp. 4052-4061. ISSN 2041-6539

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/74685/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

# Chemical Science

Accepted Manuscript



This article can be cited before page numbers have been issued, to do this please use: J. Shaw, R. Vaiyapuri, M. Parker, C. Murray, K. Lim, C. Pan, M. Knappert, C. J. Cardin, B. W. Greenland, R. Grau-Crespo and H. M. Colquhoun, *Chem. Sci.*, 2018, DOI: 10.1039/C8SC00830B.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [author guidelines](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the ethical guidelines, outlined in our [author and reviewer resource centre](#), still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

# Elements of fractal geometry in the $^1\text{H}$ NMR spectrum of a copolymer intercalation-complex: Identification of the underlying Cantor set

John S. Shaw, Rajendran Vaiyapuri, Matthew P. Parker, Claire A. Murray, Kate J. C. Lim, Cong Pan, Marcus Knappert, Christine J. Cardin, Barnaby W. Greenland, Ricardo Grau-Crespo and Howard M. Colquhoun\*

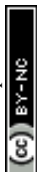
Department of Chemistry, University of Reading, Whiteknights, Reading, RG6 6AD, UK

**ABSTRACT:** Sequence-selective intercalation of pyrene into the chain-folds of a random, binary copolyimide under fast-exchange conditions results in the development of self-similar structure in the diimide region of the  $^1\text{H}$  NMR spectrum. The resulting spectrum can be described by the mathematics of fractals, an approach that is rationalised in terms of a dynamic summation of ring-current shielding effects produced by pyrene molecules intercalating into the chain at progressively greater distances from each "observed" diimide residue. The underlying set of all such summations is found to be a defined mathematical fractal namely the fourth-quarter Cantor set, within which the observed spectrum is embedded. The pattern of resonances predicted by a geometric construction of the fourth-quarter Cantor set agrees well with the observed spectrum.

## INTRODUCTION

The principle that digital information can be encoded at the molecular level, as a linear sequence of monomer residues in a high molecular weight copolymer chain, has provided a definitive basis for the whole of biology.<sup>1,2,3,4</sup> Life, however, remains the sole example of an operational information technology working at the molecular level, despite the fact that even the simplest two-monomer linear copolymer is the logical equivalent of a binary string.<sup>5</sup> Nevertheless, significant efforts are now being made to develop sequence-defined, synthetic copolymers and to explore their potential for information storage and processing.<sup>6,7,8,9,10,11,12</sup>

Interestingly, even random copolymers have potential in this field, since it is well established that the information capacity of a linear sequence reaches a maximum when the sequence is fully random.<sup>13</sup> This slightly counter-intuitive result makes sense when it is realised that "information" comes in two forms, (i) *potential* information (or information capacity), which is simply the



number of digits present when information is expressed in the form of a binary string (ignoring any possible meaning), and (ii) *contextual* information, whereby sequences within such a string acquire meaning through the application of an external context.<sup>14</sup> For example, the string of letters that comprise the title of this paper is, if not completely random, at least completely arbitrary – it only conveys information because short sequences within the string such as "fractal", "spectrum", and "intercalation" have previously been assigned meaning in the context of the English language. Indeed the letter-sequence "fractal" was given meaning only a few decades ago, by Mandelbrot;<sup>15</sup> a good example of how a specific but meaningless sequence can become information through the application of context. A possible approach to generating information from an extended but random copolymer chain would therefore be to identify short, specific sequences within the copolymer, to which meaning may eventually be assigned, then to read out these sequences by selective interaction with some form of "reader-molecule", and finally to give the identified sequences contextual meaning. Generating information by selecting sequences from a pre-existing copolymer chain is just as valid as synthesising the same sequences from scratch. Conceptually, it is the equivalent of selecting words from a dictionary rather than spelling them out letter by letter.

With regard to molecular-level reading of sequences in *synthetic* copolymers, some progress has been made using the ability of  $\pi$ -electron-rich tweezer-molecules<sup>16</sup> to bind reversibly to specific triplet sequences in chain-folding aromatic copolyimides.<sup>17,18,19,20</sup> This form of sequence-recognition involves sterically and electronically complementary  $\pi$ - $\pi$ -stacking between the tweezer arms and the diimide residues of the copolymer, producing large ring-current shielding effects,<sup>21,22</sup> and thereby resolving extended sequence-information in the <sup>1</sup>H NMR spectrum.

Here we report the results of a different form of supramolecular sequence-reading, based on selective but dynamic intercalation of the simple aromatic probe-molecule pyrene into specific chain-folds of a novel copolyimide, **1**, (Figure 1). This type of assembly reveals progressively more extended sequence-information in the <sup>1</sup>H NMR spectrum of the copolymer as the concentration of probe molecule increases. Wholly unexpectedly, the sequence-information being "read" is ultimately expressed in the form of a *self-similar pattern* in the diimide region of the spectrum. Self-similarity is perhaps the most general characteristic of objects showing fractal



geometry,<sup>23,24,25</sup> and mathematical or "true" fractals show perfect self-similarity over an infinite number of length scales. Real objects, however, are much less perfectly structured and are generally described as fractal "if they contain parts that at two or more smaller scales appear in some way similar to the whole".<sup>26</sup> There have been several reports of molecular and supramolecular structures that show self-similarity and approximate fractal geometry,<sup>27,28,29</sup> most recently via the assembly, through halogen bonding, of brominated oligophenylenes on the Ag(111) surface to give some very realistic Sierpiński triangles as visualised by STM.<sup>30</sup> In the present work however, self-similarity emerges not in the structure of a molecular assembly itself, but in the *induced magnetic field* (and consequent NMR spectrum) created by dynamic, sequence-specific intercalation of small aromatic molecules into a random copolymer chain.

## RESULTS AND DISCUSSION

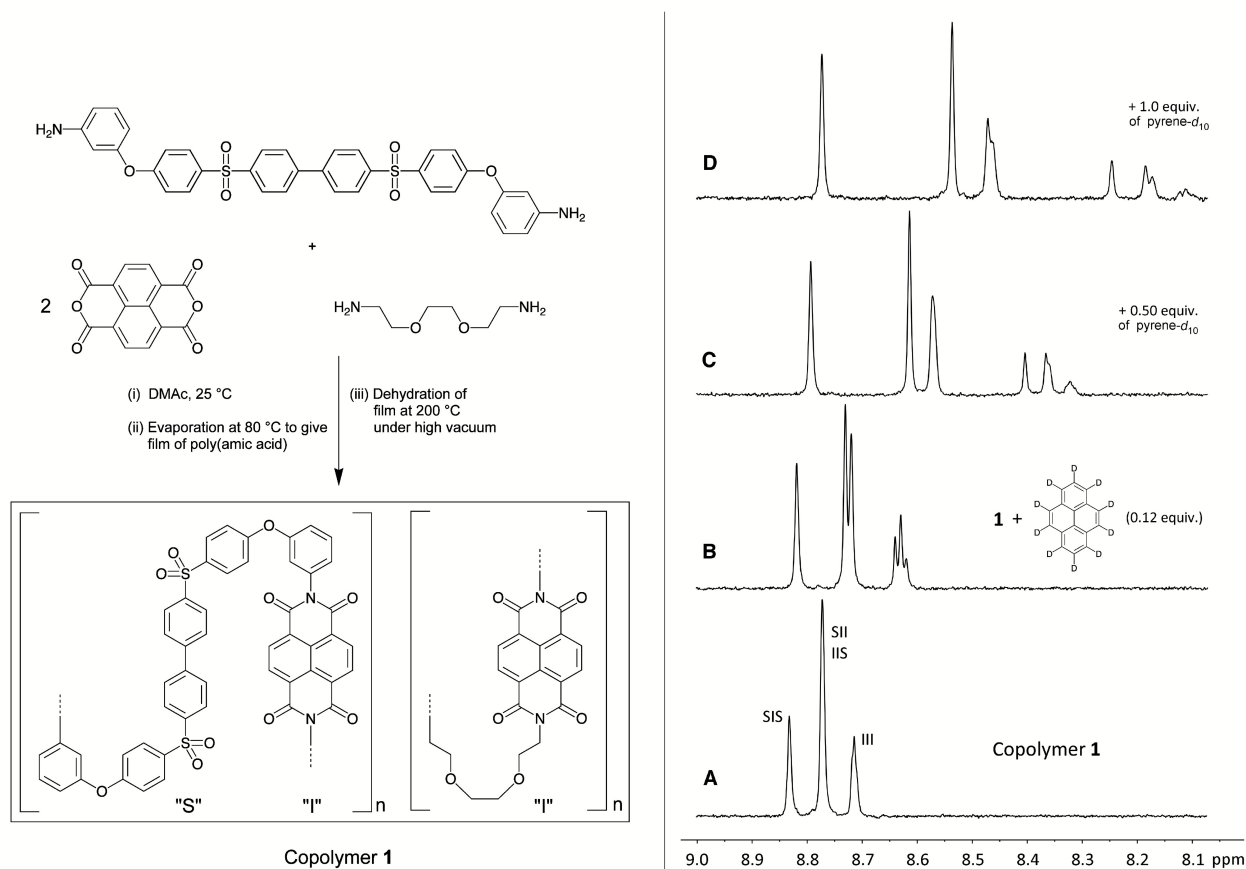
In the absence of pyrene, copolymer **1** exhibits a three-line group of resonances (relative intensities 1:2:1) at lowest field (Figure 1A). These arise from naphthalenedimide (NDI) residues flanked by either two aromatic diamine residues (Ar-NDI-Ar), two aliphatic diamine residues (Al-NDI-Al), or one of each (Ar-NDI-Al). The latter, unsymmetrical sequence is directionally degenerate, i.e. there also exists a reverse sequence (Al-NDI-Ar) which, when read backwards, is the same as the original. Thus, in a random 1:1 copolymer there are twice as many sequences of this type as there are of each of the other two sequences. The observed pattern of intensities (1:2:1) confirms that the monomer sequence, at least at the triplet level, is genuinely random.

Addition of progressively increasing amounts of pyrene (in the perdeuterated form, to avoid resonance-overlap) to a 4 mM solution of copolymer **1** produced marked splittings and large upfield shifts ( $\Delta\delta$  up to 0.62 ppm) of some of the diimide resonances (Figure 1, A-D). These effects clearly result from the binding of pyrene-*d*<sub>10</sub> molecules to the copolymer chain, and the absence of separate resonances corresponding to bound and unbound polymer shows that the system is dynamic, with fast exchange on the NMR timescale.

The most remarkable feature of the diimide region of the spectrum at high pyrene loadings (Figure 1D) is its self-similarity, as illustrated in Figure 2 where the original spectrum (a) is shown re-scaled by 1/4, 1/16, and 1/64. In each case, the re-scaled spectrum closely resembles

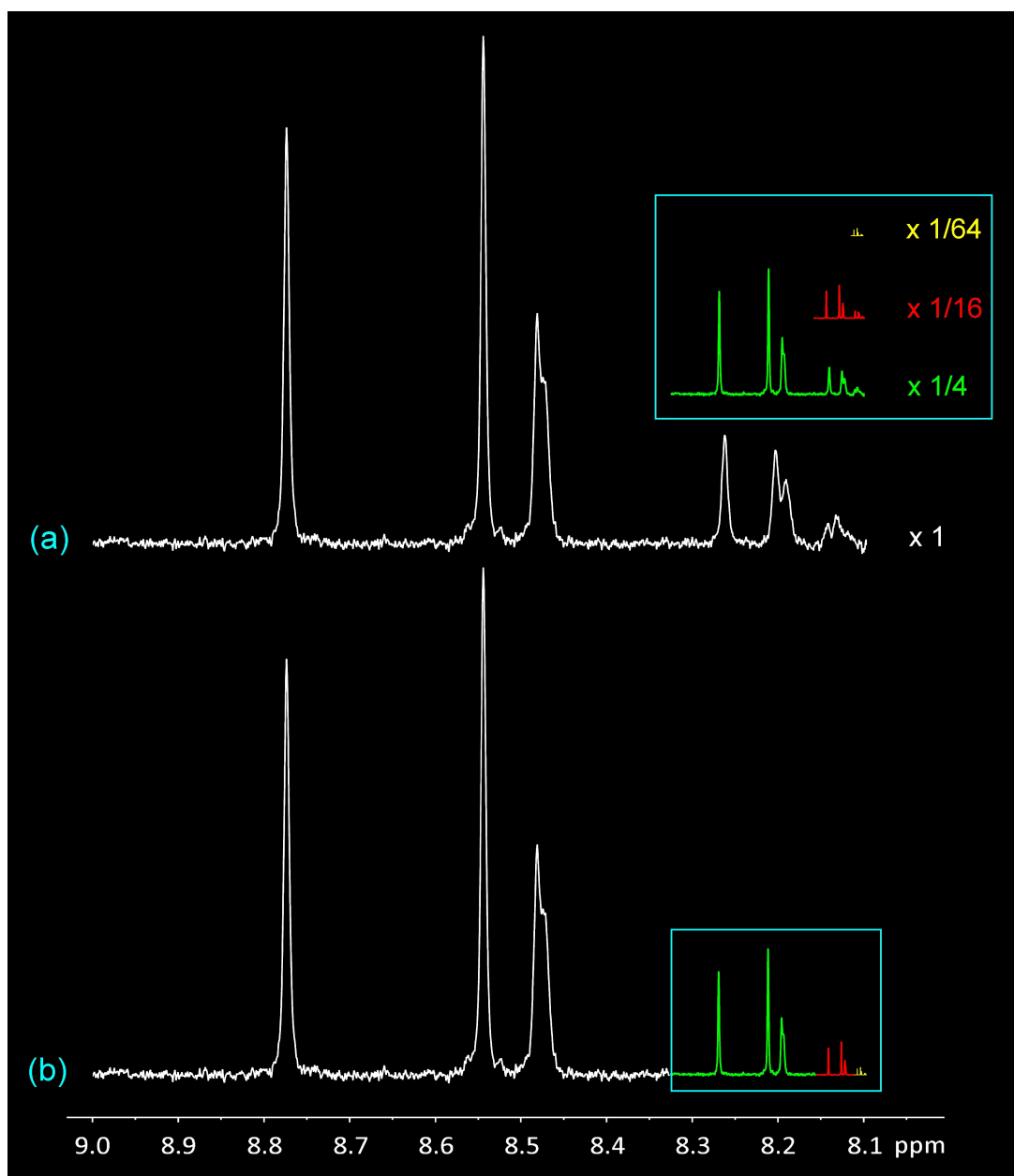


the highest-field quarter of the previous-scale spectrum. Spectrum (b) is a geometric reconstruction in which these re-scaled copies successively replace the highest field 1/4 of the previous spectrum. Even allowing for the effects of scaling on spectroscopic linewidths, it is clear from the reconstruction shown in Figure 2(b) that the experimental spectrum does indeed consist – to good approximation – of smaller copies of itself over at least three length scales, and can therefore be described as self-similar.<sup>26</sup>

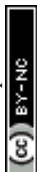


**Figure 1.** Synthesis of binary copolymer **1** and the <sup>1</sup>H NMR spectrum (diimide region) of this copolymer at 25 °C in the presence of increasing concentrations of pyrene-*d*<sub>10</sub>. **A:** Pure copolymer **1** (in CDCl<sub>3</sub>/(CF<sub>3</sub>)<sub>2</sub>CHOH, 6:1 v/v) at 4 mM concentration of NDI residues. **B, C, D:** Copolymer **1** in the presence of 0.12, 0.50 and 1.0 molar equivalents (relative to NDI residues) respectively of pyrene-*d*<sub>10</sub>.





**Figure 2.** Self-similarity in the  $^1\text{H}$  NMR spectrum of copolymer **1** (diimide region) in the presence of one equivalent of pyrene- $d_{10}$ . **(a):** Spectrum "D" from Figure 1, expanded vertically. The inset shows the *entire* spectrum re-scaled  $\times 1/4$  (green),  $\times 1/16$  (red) and  $\times 1/64$  (yellow) in both dimensions. In every case the rescaled spectrum closely resembles the highest field one-quarter of the spectrum at the previous length scale. **(b):** Reconstruction of spectrum "(a)" by deleting the highest field 1/4 of the spectrum and replacing it with a 1/4 scale copy of the original spectrum, and then repeating the process over two further length scales. Although this type of geometric reconstruction does not reproduce the original spectrum exactly, mainly owing to progressive narrowing of the spectroscopic linewidth, comparison of (a) and (b) brings out the self-similar character of spectrum (a) very clearly.





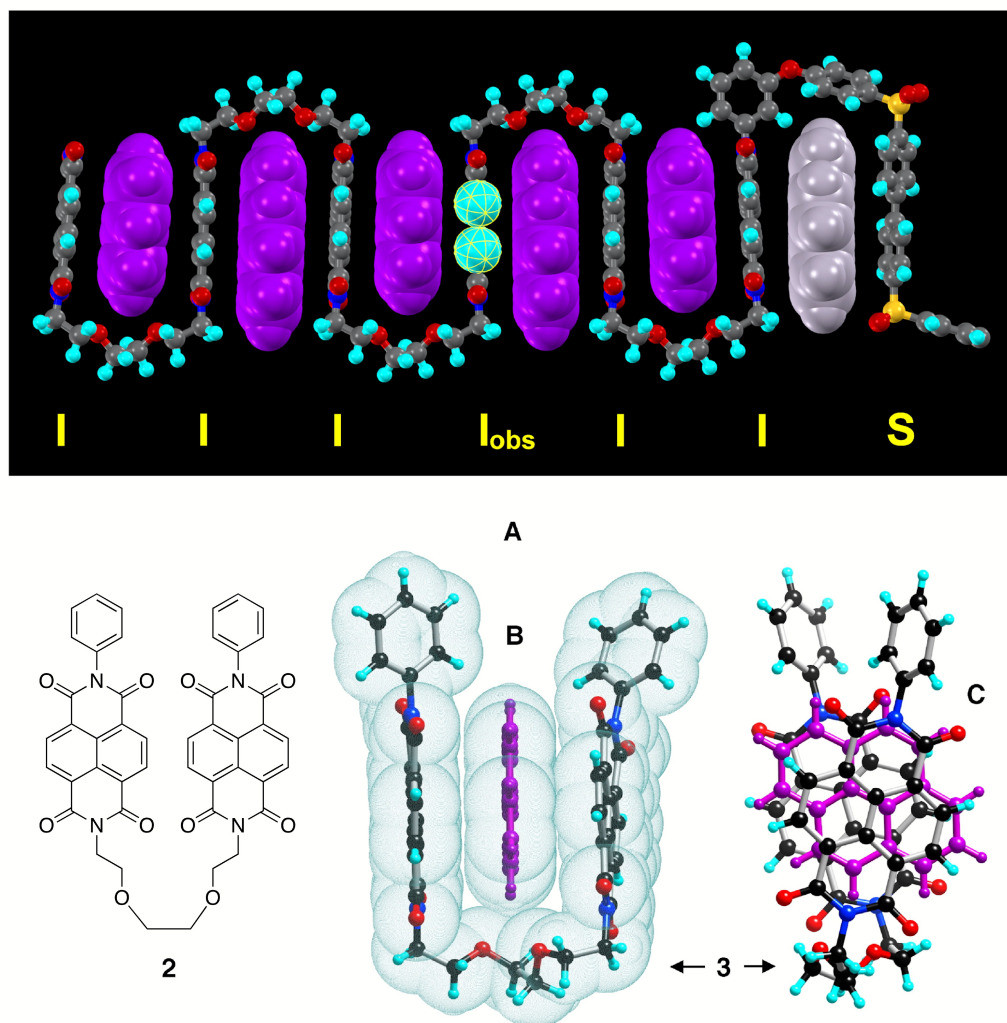
We have developed an atomistically well-defined model that predicts the emergence of self-similar character in the  $^1\text{H}$  NMR spectrum of the pyrene-intercalated copolymer, and even allows the underlying mathematical fractal to be identified. This model builds on the idea that the total ring-current shielding experienced by a diimide proton is the sum of the diminishing shielding effects produced by intercalator-molecules that bind at progressively greater distances along the polymer chain from a central, "observed", diimide residue (Figure 3A).<sup>17,18,19,20</sup> Defining a chain-fold comprising two NDI units linked by a 1,2-bis(2-aminoethoxy)ethane residue as "II" and one comprising an NDI unit adjacent to a 4,4'-biphenylenedisulfone group as "IS" (or "SI"), then our initial model is illustrated in Figure 3A as an energy-minimised, chain-folded and intercalated septet sequence.

However, this initial model takes no account of differences in binding constant between pyrene and the different chain-folds present in copolymer **1**. Figure 1 shows clearly that there must be large differences of this type, because resonances representing different sequences are shifted upfield by anything from 0.07 ppm (for "SIS"-centred sequences) to 0.62 ppm (for "III"-centred sequences). [Note that the sequence "SS" is forbidden by the chemistry of the copolyimide, since the diamine residue "S" must always be followed by (and preceded by) the diimide residue "I"].

Crystallographic evidence for strong complexation of pyrene at the "II" chain-fold was provided by single crystal X-ray analyses of the model complex **3** (Figure 3B). Here, a pyrene molecule is bound by an oligomer (**2**) comprising two NDI residues linked by a single fold-unit and terminated by phenyl residues. The area of overlap between the pyrene and NDI units is very high (Figure 3C), and their separations (atom-to-plane distances of 3.3 - 3.5 Å ) correspond to tight van der Waals contact (Figure 3B). Moreover, the recently-reported crystal structure of a different complex, now between oligomer **2** and a bis-pyrenyl tweezer molecule provides a very good model for the multiply-intercalated structure proposed in Figure 3A.<sup>31</sup> See Supporting Information.





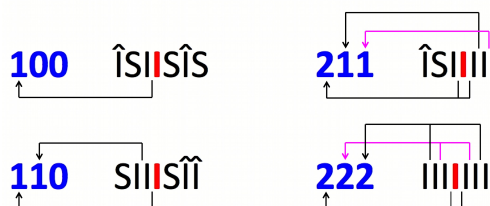


**Figure 3.** (A). Energy-minimised model for a segment of copolymer **1** (septet sequence "IIIIIS") intercalated by pyrene. Diimide residues are indicated as "I" and the biphenylenedisulfone residue as "S". The central, "observed" diimide residue is indicated as "I<sub>obs</sub>". Pyrene molecules located in the strongly-binding "II" chain-fold are shown in purple and a pyrene molecule located in the very weakly-binding "IS" fold is shown in grey. (B). Single crystal X-ray structure of a chain-folded 1:1 complex (**3**) formed between dimer **2** and pyrene viewed along the plane of the intercalating pyrene molecule. The fold provides a binding site in which the measured perpendicular distance between the centroids of the two diimide residues is 6.82 Å. (C). Complex **3** viewed perpendicular to the plane of the pyrene molecule (shown in purple).

Given the evidently large difference in binding constant between sequences "SI" and "II", we next adopt the simplifying approximation that pyrene binds *only* to "II" sequences so that, in a predictive binding model, the positions of "II" pairs are the defining characteristics of any longer



sequence. For example, reading outwards in both directions from the central "observed" diimide residue (highlighted here in red), the septet sequence ISIIII contains *two* "II" sites immediately adjacent to – and including – the central residue. In addition, there is *one* "II" site at a next-adjacent position, and *one* "II" site at an outermost position. In binding terms, the sequence ISIIII can therefore be described by the 3-digit code [211]. See Figure 4.



**Figure 4.** Methodology for assigning 3-digit codes to different septet sequences in copolymer **1**. For each septet sequence, the number of "II" pairs at the three types of position relative to the central, observed diimide residue **I** (adjacent, next-adjacent and next-next-adjacent) gives the three digits of the corresponding code. Note that diimide residues outside an "S" residue (relative to the centre) are coded as  $\hat{I}$ . These do not contribute to the total ring current shielding of **I** for reasons discussed below.

Any imide-centred septet sequence can similarly be assigned a 3-digit code in which the highest possible value of any digit is 2 and the lowest value is zero since, in any linear copolymer the number of monomer residues of a given type at a specific distance from any other monomer residue can only be 0, 1 or 2. The code for IIIIII would then be [222] and, for SIIIIIS, [220]. The experimental data, however, reveal an additional requirement for translating copolymer sequences into digital codes based on "II" pairs. This is that pyrene intercalation into "II" binding sites *outside* an "S" residue (relative to the central "I" residue) must be ignored. This effect may result from the weakly binding "SI" sequence allowing a gross departure from the time-averaged, chain-folded conformation (Figure 3A) that enables long range ring-current shielding of the central "I" by intercalating pyrene molecules. Examples of this methodology for assigning digital codes to copolymer sequences are illustrated in Figure 4, where "I" residues outside an S residue are given the symbol " $\hat{I}$ ".

The 27 three-digit codes containing only the digits 0, 1 or 2 are shown in Table 1. Although every possible septet-sequence in copolymer **1** can be assigned a code, not every code corresponds to a realisable sequence, reducing the number of assignable codes from twenty-seven possibilities to just the ten codes shown in Table 1. A number of codes show redundancy



in that, from the rules of the model described above, they can represent two or even three of the 15 different sequences. There are 64 ( $2^6$ ) "I"-centred septet combinations of "I" and "S", but 39 of these contain "SS" pairs which are not allowed by the chemistry of the system. This reduces the total to 25, but 10 of these are unsymmetrical and so are indistinguishable in NMR terms from their reverse sequences, reducing the total to just the 15 sequences shown in Table 1.

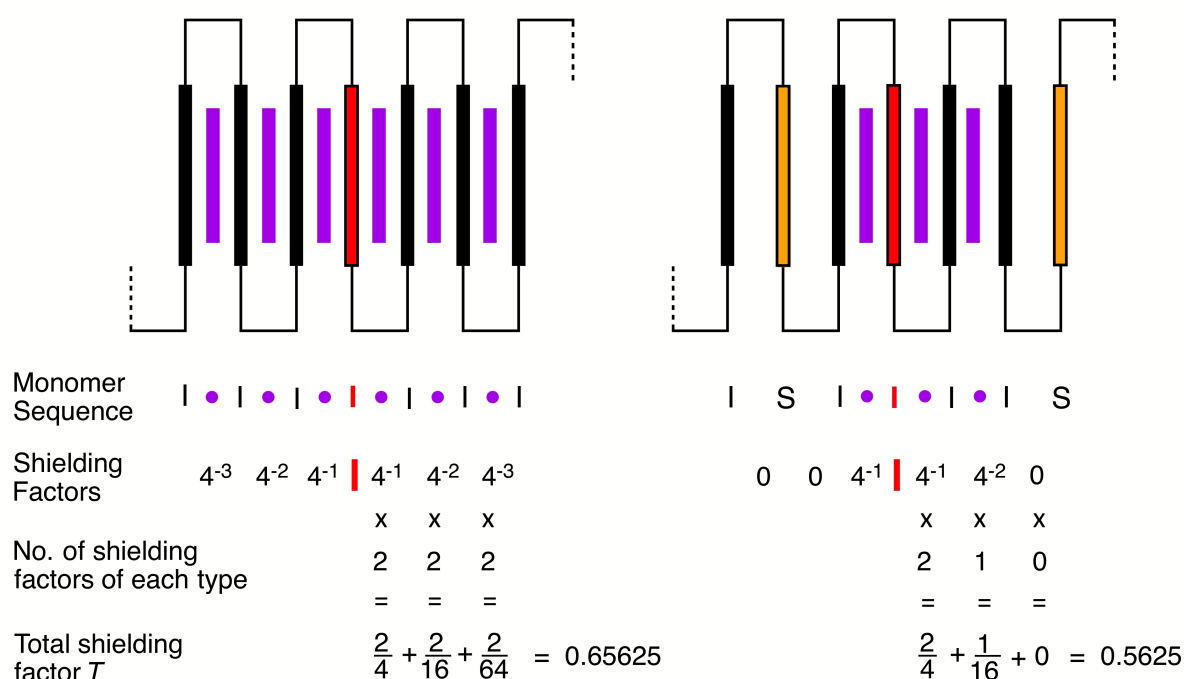
**Table 1.** The 27 three-digit combinations of 0, 1 and 2, and their associated I-centred septet sequences (where allowed) in copolymer **1**. Codes arising from allowed sequences in copolymer **1** are shown bold in blue. The observed diimide residue at the centre of each allowed sequence is highlighted as **I**, and neighbouring diimide residues are shown as either "I" or  $\hat{I}$  depending on whether they are "inside" or "outside" an S residue (relative to the central **I**). Pyrene binding to  $\hat{I}\hat{I}$  pairs appears to produce no additional shielding of the central **I** residue, and such pairs are therefore ignored when assigning digital codes to sequences. The unused codes in this Table correspond to patterns of "II" pairs for which no septet sequences can exist in copolymer **1**.

	<b>0</b>	<b>1</b>	<b>2</b>	
<b>0</b>	<b>000</b> $\hat{S}\hat{I}\hat{S}\hat{I}\hat{S}$ , $\hat{S}\hat{I}\hat{S}\hat{I}\hat{S}\hat{I}$ , $\hat{I}\hat{I}\hat{S}\hat{I}\hat{S}\hat{I}$ 001 002	010 011 012	020 021 022	<b>0</b> <b>1</b> <b>2</b>
<b>1</b>	<b>100</b> $\hat{I}\hat{S}\hat{I}\hat{I}\hat{S}\hat{I}$ , $\hat{I}\hat{S}\hat{I}\hat{I}\hat{S}\hat{I}$ 101 102	<b>110</b> $S\hat{I}\hat{I}\hat{S}\hat{I}\hat{S}$ , $S\hat{I}\hat{I}\hat{S}\hat{I}\hat{S}\hat{I}$ <b>111</b> $\hat{I}\hat{I}\hat{I}\hat{S}\hat{I}\hat{S}$ , $\hat{I}\hat{I}\hat{I}\hat{S}\hat{I}\hat{S}\hat{I}$ 112	120 121 122	<b>0</b> <b>1</b> <b>2</b>
<b>2</b>	<b>200</b> $\hat{I}\hat{S}\hat{I}\hat{I}\hat{I}\hat{S}$ 201 202	<b>210</b> $\hat{I}\hat{S}\hat{I}\hat{I}\hat{I}\hat{S}$ <b>211</b> $\hat{I}\hat{S}\hat{I}\hat{I}\hat{I}\hat{I}$ 212	<b>220</b> $S\hat{I}\hat{I}\hat{I}\hat{I}\hat{S}$ <b>221</b> $S\hat{I}\hat{I}\hat{I}\hat{I}\hat{I}$ <b>222</b> $\hat{I}\hat{I}\hat{I}\hat{I}\hat{I}\hat{I}$	<b>0</b> <b>1</b> <b>2</b>

The magnitude of shielding of the central, "observed" diimide residue produced by each intercalating pyrene molecule diminishes progressively (we estimate, empirically, by a factor of about four per chain-fold under the conditions of the experiment) as it binds further and further



out from the centre of the sequence. The sequence IIIIII, described by the code [222], thus produces the maximum possible degree of ring-current shielding of the central diimide residue when the sequence is intercalated by pyrene. The sequence ISIIIS, described by the code [210], will lead to a lesser degree of cumulative shielding, and the effect can be quantified as illustrated in Figure 5. The integers of the digital shielding code can be shown to represent the coefficients of a quaternary expansion of a total (decimal) shielding parameter  $T$  for that sequence. In the presence of pyrene, sequences with higher values of  $T$  will produce larger complexation shifts of the central diimide resonance than sequences with lower values of  $T$ .



**Figure 5.** Cumulative magnetic shielding effects produced by pyrene molecules (purple) intercalating into copolymer chain-folds (black for "I" and gold for "S") that are progressively more distant from a central, "observed" diimide residue. Two examples of septet-sequences in copolymer **1** are shown, along with their intercalated pyrene molecules. The large difference in binding constant for pyrene between the sequences "II" and "IS" is approximated by ignoring binding to "IS" (and "SI") altogether. Ring-current shielding of the central diimide protons by intercalating pyrene molecules is proposed to fall off by a factor of approximately four at each successively more distant chain fold. Each different copolymer septet-sequence gives rise to a three-figure code representing the number of pyrene molecules (0, 1 or 2) that can intercalate at each of the three types of position in the sequence. The total magnetic shielding,  $T$ , experienced by the central diimide protons of any septet sequence is the sum of the individual shielding factors, taking account of their multiplicity (0, 1 or 2).



Figure 5 shows the summation of shielding factors for just two possible septet sequences, but the approach is valid for all allowed septets, so that in general terms:

$$T = \sum_{k=1}^3 \frac{N_k}{4^k} \quad (\text{Equation 1})$$

where  $N_k$  takes the values 0, 1 or 2, as defined in Table 1. These are the only possible values of  $N_k$ , because, in a linear copolymer, the number of monomer residues of a given type at a specific distance from any other monomer residue can only be 0, 1 or 2. Equation 1 thus represents a rather general description of the intramolecular environment of the central monomer residue in any binary copolymer sequence. There is, moreover, no requirement to limit the sequence-length to septets: small but incremental increases in the total shielding factor will result from pyrene intercalation at even quite distant chain folds.

Then, for an *infinite* "I"-centred sequence in copolymer **1** we would have:

$$T = \sum_{k=1}^{\infty} \frac{N_k}{4^k} \quad (\text{Equation 2})$$

where, again,  $N_k$  may (only) take the values 0, 1 or 2. The proposed exponential fall-off in shielding with distance clearly differs from the quadratic decline found for ring-current shielding of a single static proton within the shielding zone of a single aromatic molecule.<sup>22</sup> However, the present dynamic system, with its interpolating pyrene molecules and diimide units between the shielding molecules and the observed protons, obviously cannot be represented by such a simple model. As show below, the exponential model leads to excellent agreement with experiment, indicating that the fall-off is indeed exponential in this system. The value of 4 for the exponential base is empirical and is determined by the NMR data. In the Supporting Information we show that the exponential base does not need to be an integer (in fact, direct fitting of the model against experimental data leads to a value of  $3.9 \pm 0.1$ ), but we refer to its value as approximately four in order to introduce more easily the elegant geometric representation given below.

Equation 2, which emerges from purely structural and magnetic considerations, defines a known, specific mathematical fractal which we term the "fourth-quarter" Cantor set. This fractal was first described (in more general terms) by Smith in 1875,<sup>32</sup> but so far as we are aware it has not been



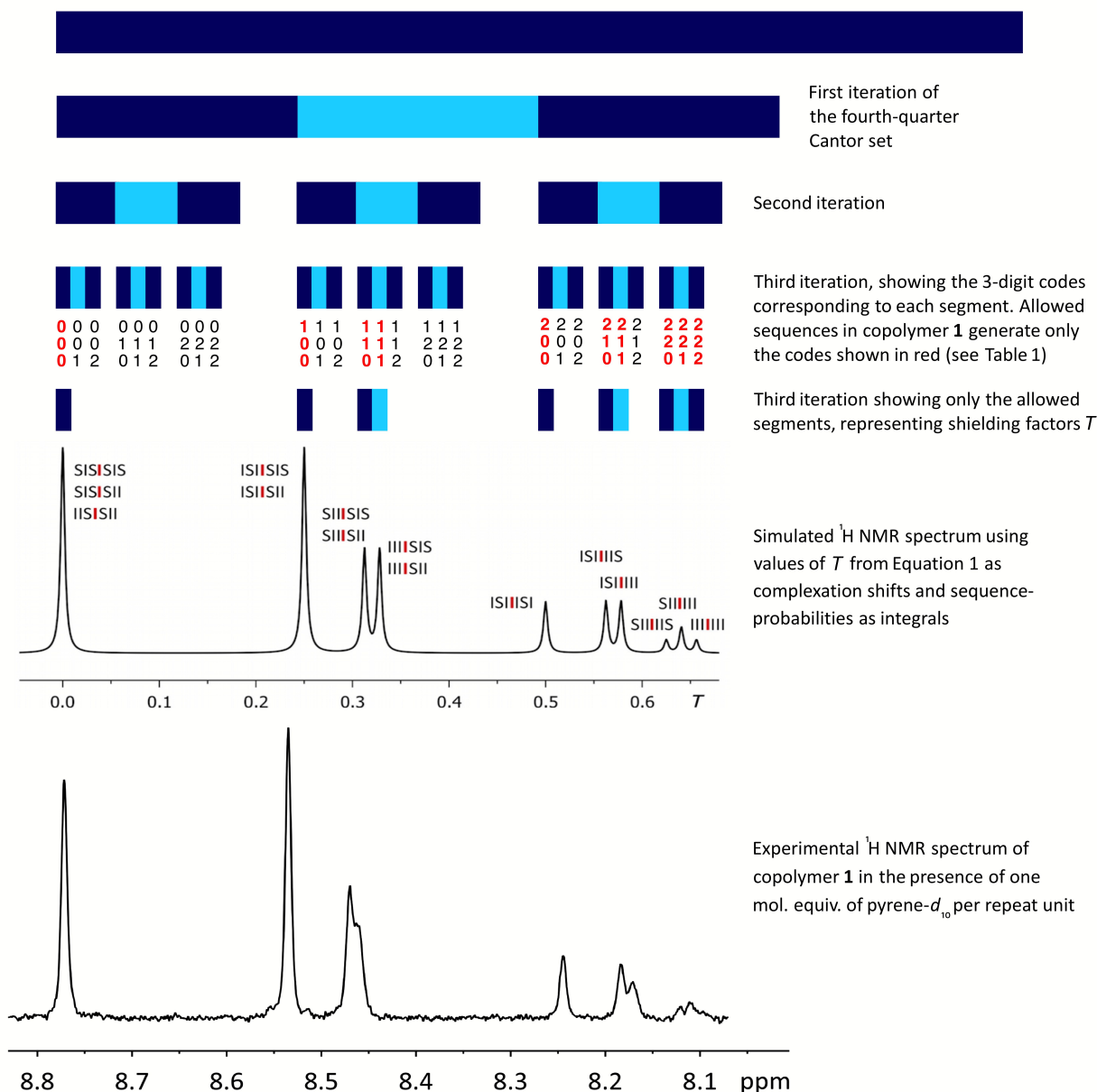
found to underlie phenomena in the physical world until now. It is a variant of the more famous "middle-third" Cantor set,<sup>33,34,35</sup> which may be defined as "those numbers in the interval [0, 1] that admit ternary expansions not containing the integer 1".<sup>23</sup> Geometrically, the latter set is constructed by dividing a line of unit length into three equal segments; deleting the middle segment; and repeating these operations on the remaining segments. What remains after an infinite number of such divisions and deletions is the middle-third Cantor set.<sup>25</sup>

Since the values of  $N_k$  in Equation 2 represent the coefficients of a *quaternary* expansion, and the possible integers of a quaternary number are 0, 1, 2 and 3, then the **fourth-quarter** Cantor set (Equation 2) can be defined as "those numbers in the interval [0, 1] that admit quaternary expansions not containing the integer 3". In the Supporting Information we include a formal proof of the equivalence between Equation 2 and the fourth-quarter Cantor set. The geometric construction of this set (Figure 6) involves dividing a line of unit length into four equal segments, deleting the fourth quarter, repeating these operations on the remaining segments, and so on indefinitely. The set converges to an upper limit of 0.666....., and has a fractal (Hausdorff) dimension of 0.792. Unlike the middle-third set, the "remaining segments" of the fourth-quarter set are contiguous and so, in Figure 6, are coloured to allow the construction to be visualised.

Each iteration increases the sequence-length represented in the model, with the first iteration describing, in the present system, the induced magnetic field arising from triplet sequences, the second from quintets, the third from septets, and so on without limit. After three iterations the set contains 27 segments corresponding to the 27 *septet-based* codes shown in Table 1. If we now delete the segments corresponding to codes which cannot be generated by any permitted sequence in copolymer **1**, a pattern of shielding factors emerges (Figure 6; codes shown in red) that corresponds very closely to the pattern of complexation shifts observed in the NMR experiment. In the final model, complexation shifts,  $\Delta\delta$ , displayed by the central diimide resonance in any given monomer sequence are proportional to the total shielding factor  $T$  for that sequence. As the proportionality constant between  $\Delta\delta$  and  $T$  is very close to 1 ppm, we can set  $T$  (a dimensionless number) *equal* to  $\Delta\delta$  in ppm, thus enabling simulation of a  $^1\text{H}$  NMR spectrum in which the complexation shift of each septet is given by Equation 1, using as coefficients ( $N_k$ ) the digits of the codes shown in Table 1.

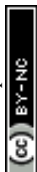






**Fig. 6.** First three iterations of the geometrical construction of the fourth-quarter Cantor set and comparison of the  $^1\text{H}$  NMR spectrum predicted from this set with the observed spectrum of copolymer **1** in the presence of one molar equivalent of pyrene- $d_{10}$ . The NMR simulation is based on all physically-permitted **I**-centred septet sequences present in the copolymer (Table 1), and so allows assignment of all ten resonances to the septets shown above.

Of course resonance *intensities* (integrals) are also required for any meaningful  $^1\text{H}$  NMR simulation, and in a random copolymer these are given simply by the relative probabilities of all possible septet sequences. Values of  $T$  for each allowed septet, calculated from Equation 1, together with the corresponding sequence-probabilities, are shown in Table 2. No model is

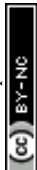




required to account for these sequence-probabilities: in a hypothetical, infinite, random copolymer chain they are rigorously defined, and a  $^1\text{H}$  NMR sample containing trillions of high molecular weight copolymer chains is a close approximation to such a chain. It is interesting that the set of sequence-probabilities for copolymer **1** shows the same type of self-similarity (a fourfold contracting transformation) as that found for the set of shielding factors  $T$ . Thus, in the diimide region, the *entire* NMR spectrum – in terms of both chemical shifts and resonance intensities – is self-similar. The predicted values of total shielding  $T$  (from Equation 1) and sequence probabilities  $P$  are shown in Table 2.

**Table 2.** Total shielding factors  $T$  and sequence-probabilities  $P$  for the allowed **septet** sequences in copolymer **1**. The simulated  $^1\text{H}$  NMR spectrum shown in Figure 6 was generated by plotting "Group probability" (final column) against "Group  $T$ -value" (fourth column).

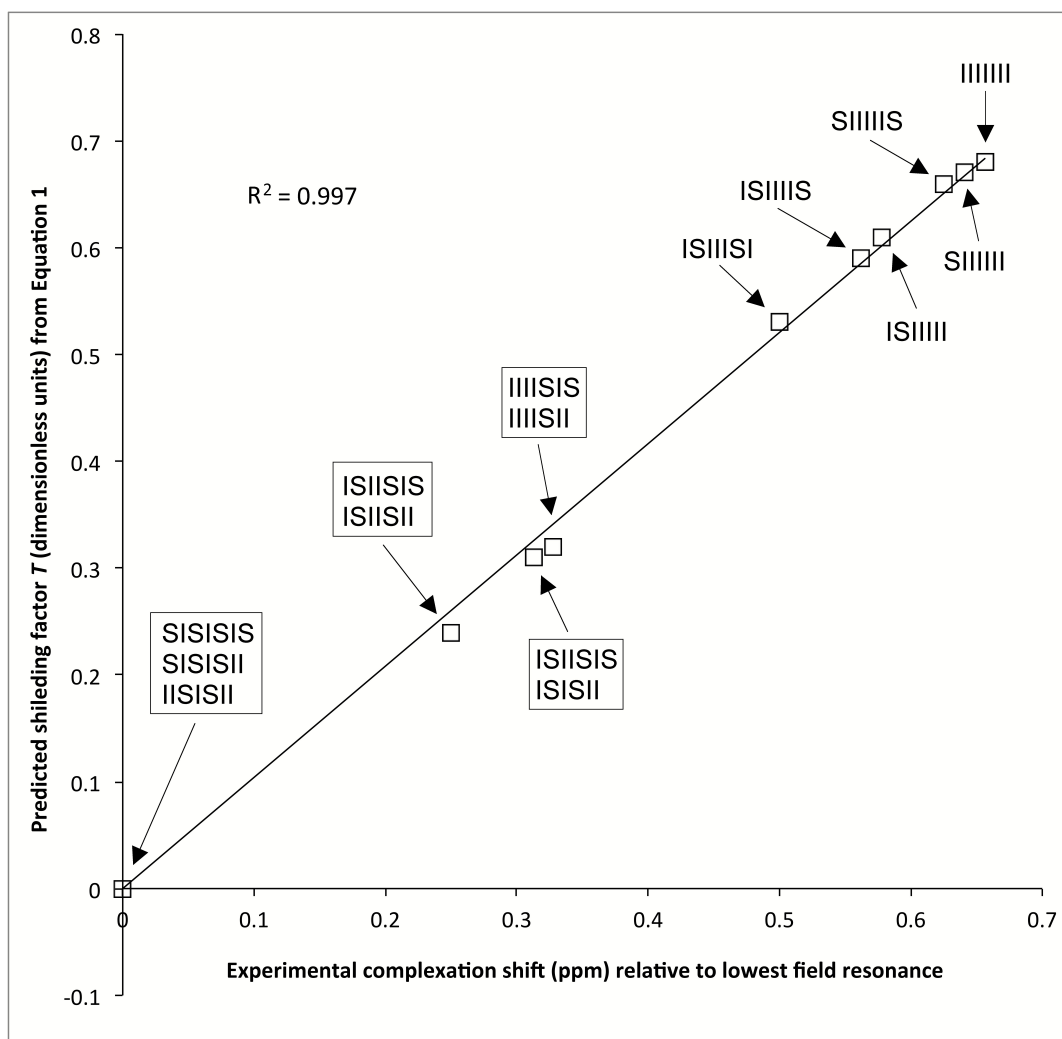
Septet sequence	Digital code	Shielding factor $T (= \Delta\delta)$	Group $T$ -value $(= \Delta\delta)$	Sequence probability $(P)$	Directional degeneracy $(D)$	Overall probability $(P \times D)$	Group probability $[\Sigma(P \times D)]$
SISIS	000	0		0.0625	1	0.0625	
SISISII	000	0		0.0625	2	0.125	
IISISII	000	0	0	0.0625	1	0.0625	0.25
ISISIS	100	0.25		0.0625	2	0.125	
ISISII	100	0.25	0.25	0.0625	2	0.125	0.25
SIIISIS	110	0.3125		0.03125	2	0.0625	
SIIISII	110	0.3125	0.3125	0.03125	2	0.0625	0.125
IIISIS	111	0.328125		0.03125	2	0.0625	
IIISII	111	0.328125	0.328125	0.03125	2	0.0625	0.125
ISIIISI	200	0.5	0.5	0.0625	1	0.0625	0.0625
ISIIIS	210	0.5625	0.5625	0.03125	2	0.0625	0.0625
ISIIII	211	0.578125	0.578125	0.03125	2	0.0625	0.0625
SIIIIIS	220	0.625	0.625	0.015625	1	0.015625	0.015625
SIIIII	221	0.640625	0.640625	0.015625	2	0.03125	0.03125
IIIIII	222	0.65625	0.65625	0.015625	1	0.015625	0.015625



In the present context we should emphasise the distinction between fractality and self-similarity. The self-similar character of our experimental NMR spectrum (Figure 2) does not mean that the spectrum itself can be described as a fractal. What *is* a fractal is the underlying set of all shielding factors, given by Equation 1, within which the physically-permitted contributions are embedded. As seen in Figure 2, the spectrum shows scaling about just a single point (at ca. 8.1 ppm), so that the pattern contains only one copy of itself at each scale. Mathematical or "true" fractals show scaling about *several* points: the classical "middle-third" Cantor set, for example, contains copies of itself in the first and last third of the set, and the fourth-quarter Cantor set described here contains copies of itself in the first, second and third quarters. The self-similar characteristics of the observed  $^1\text{H}$  NMR spectrum (Figure 3) therefore represent only elements of the latter fractal, and the self-similarity of this spectrum is equivalent to that of the so-called logarithmic spiral, whose structure appears the same when magnified about just a single point (the origin).<sup>36</sup> It may be noted that logarithmic spirals are found embedded within several fractals, including the well-known Mandelbrot set,<sup>15</sup> in much the same way that the NMR spectrum under discussion here is embedded within the fourth-quarter Cantor set.

As noted above, Figure 6 shows a simulated  $^1\text{H}$  NMR spectrum generated using total shielding factors  $T$  to predict chemical shifts and sequence-probabilities  $P$  to give predicted integrals. Correlation of the predicted spectrum with the experimental spectrum of the pyrene complex of copolymer **1** is remarkably good. Indeed, a correlation plot of  $T$  vs  $\Delta\delta$  is very close to being linear, as shown in Figure 7. Moreover, Equation 2 predicts that sequences longer than septets, represented by further iterations of the fractal construction, should provide even more realistic predictions of shielding factors and complexation shifts, and this is borne out by simulation of a spectrum now based on nonet sequences (see Supporting Information).

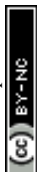




**Figure 7.** Values of  $T$  (dimensionless units) calculated from Equation 1 plotted against observed complexation shifts for the 1:1 copolymer-pyrene complex ( $\Delta\delta$ , ppm), relative to the lowest field resonance as zero. This choice of zero removes any contribution from internal, non-complexation-related shielding within the copolymer chain. The close fit to linearity provides good evidence for the validity of the shielding model proposed in Figures 5 and 6.

## CONCLUSIONS

This work describes a copolymer intercalation-complex whose  $^1\text{H}$  NMR spectrum contains almost exact copies of itself over several different length scales, demonstrating for the first time the existence of self-similarity in an NMR spectrum. The geometrical structure of the spectrum is rationalised on the basis of an atomistically-defined, supramolecular model that predicts a direct relationship between the observed spectrum and the mathematics of fractals, specifically the



fourth-quarter Cantor set. A simulation of the spectrum based on this set provides a remarkably good match to the experimental data and at the same time enables all the observed resonances to be assigned to specific sequences or groups of sequences. This type of fractal can in fact provide a general, mathematical description of the environment of any monomer residue within a binary copolymer chain, and we are currently exploring its application to a wide range of copolymer systems and their supramolecular complexes.

## EXPERIMENTAL

### General considerations

Synthetic procedures were carried out under an atmosphere of dry nitrogen unless otherwise specified. Solvents and reagents were obtained from Sigma-Aldrich and used without purification unless otherwise stated. *N,N*-Dimethylacetamide (DMAc) and *N,N*-dimethylformamide (DMF) were distilled over calcium hydride before use. 4,4'-Bis[4-(3-aminophenoxy)benzenesulfonyl]-1,1'-biphenyl was prepared according to a literature procedure,<sup>37</sup> as was the model oligomer **2**.<sup>31</sup> Proton and <sup>13</sup>C NMR spectra were recorded on a Bruker Avance 400 MHz spectrometer, with chemical shifts referenced to residual solvent resonances and reported in ppm relative to TMS. Inherent viscosities ( $\eta_{inh}$ ) were measured at 25 °C in a thermostatted water bath on 0.1% polymer solutions in 1-methylpyrrolidinone (NMP) using a Schott-Geräte CT-52 semi-automated viscometer and AVS 470 measurement system. Gel permeation chromatography was carried out using a Polymer Laboratories PL-220 instrument fitted with 2 x PL 10  $\mu$ m mixed B columns. Computational modelling using molecular mechanics with charge-equilibration, (DREIDING force-field, *Materials Studio*, v. 7.0, Accelrys Inc., San Diego) was carried out in order to obtain a preliminary structural model for copolymer-pyrene interactions. Simulated NMR spectra were generated using the "Peak Table to Spectrum" script in Mestrenova (v. 9.0, Mestrelab Research, Santiago de Compostela). Single crystal X-ray data were measured for the pyrene complex **3** on an Oxford Diffraction X-Calibur Gemini diffractometer at 150 K using Cu-K $\alpha$  radiation.

### Copolymer 1: synthesis and characterisation

*Copolymer 1*: 4,4'-Bis[4-(3-aminophenoxy)benzenesulfonyl]-1,1'-biphenyl (0.400 g,  $6.2 \times 10^{-4}$  mol) and 2,2'-(ethylenedioxy)bis(ethylamine) (0.090 g,  $6 \times 10^{-4}$  mol) were dissolved in dry



DMAc (5 mL) by stirring at room temperature. When the diamines had dissolved completely, 1,4,5,8-naphthalenetetracarboxylic dianhydride (0.330 g,  $1.2 \times 10^{-3}$  mol) was added in one portion. The suspension was sonicated for 2 h until homogeneous. The dark brown solution was then stirred overnight before being transferred to a Petri dish and heated under vacuum to 80 °C to remove the solvent. The resulting poly(amic acid) film was imidized by heating under vacuum at 150 °C for 30 min, 180 °C for 30 min and finally at 200 °C for 1 h. The film was then dissolved in DMF (5 mL) and reprecipitated from methanol to afford copolyimide **1**. Traces of a macrocyclic by-product<sup>38</sup> were then removed by Soxhlet extraction with chloroform and subsequent re-precipitation of the copolymer from DMF solution in methanol.

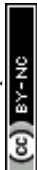
Yield 0.52 g, 74%.  $T_g = 244$  °C, inherent viscosity ( $\eta_{inh}$ ) = 0.23 dL g<sup>-1</sup>; GPC (0.1% w/w LiCl in DMF as eluent with RI detection)  $M_n = 17,800$ ,  $M_w = 27,500$  relative to polystyrene standards. <sup>1</sup>H NMR (CDCl<sub>3</sub>/CF<sub>3</sub>COOH 6:1 v/v)  $\delta$  8.93 (s), 8.88-8.82 (m), 8.79 (s), 8.06-7.99 (m), 7.75-7.67 (m), 7.65-7.58 (m), 7.24-7.17 (m), 7.13-7.04 (br), 4.54-4.44 (m), 4.03-3.94 (m), 3.92-3.82 (m); <sup>13</sup>C NMR (CDCl<sub>3</sub>/CF<sub>3</sub>COOH 6:1 v/v): 163.7, 163.5, 156.0, 144.3, 140.6, 135.3, 134.1, 132.2, 131.7, 131.4, 130.1, 128.5, 128.1, 127.0, 126.8, 126.3, 126.2, 125.0, 121.5, 120.4, 69.7, 68.0, 39.6; FT-IR  $\nu_{max}/cm^{-1}$  1709, 1668, 1581, 1485, 1341, 1247, 1153, 1107, 767.

### <sup>1</sup>H NMR titration of copolymer **1** with pyrene-*d*<sub>10</sub>.

Equimolar stock solutions with exact concentrations of pyrene-*d*<sub>10</sub> (24 mM) and copolymer **1** (24 mM) with respect to diimide residues) in CDCl<sub>3</sub>/hexafluoropropan-2-ol (6:1 v/v) were prepared by weighing an approximate amount of sample and adding the required volume of solvent via a micropipette. Copolymer solution (100  $\mu$ L) was added to each NMR tube using a micropipette and varying volumes of pyrene solution were then added to the tubes (e.g. 100  $\mu$ L for 1:1 molar ratio of pyrene to diimide). The NMR tubes were all then filled to 600  $\mu$ L total volume with the required volume of solvent (e.g. 400  $\mu$ L to the 1:1 solution) and were well-mixed, thus affording a constant diimide concentration of 4 mM in each tube.

### Crystal data for complex **3**

C<sub>46</sub>H<sub>30</sub>N<sub>4</sub>O<sub>10</sub>·C<sub>16</sub>H<sub>10</sub>·(CF<sub>3</sub>)<sub>2</sub>CHOH,  $M_r = 1169.06$ , triclinic, *P*-1,  $a = 10.6269(3)$ ,  $b = 14.9210(5)$ ,  $c = 17.4851(7)$  Å,  $\alpha = 108.065(3)$ ,  $\beta = 102.085(3)$ ,  $\gamma = 91.766(2)^\circ$ .  $V = 2563.73(11)$  Å<sup>3</sup>,  $T = 150(2)$  K,  $Z = 2$ ,  $D_c = 1.515$  g cm<sup>-3</sup>,  $\mu(\text{Cu-K}\alpha) = 1.006$  mm<sup>-1</sup>,  $F(000) = 1204$ . Independent



measured reflections 8183.  $R_1 = 0.0516$ ,  $wR_2 = 0.0927$  for 6639 independent observed reflections [ $2\theta \leq 126^\circ$ ,  $I > 2\sigma(I)$ ]. CCDC 1503298.

## SUPPORTING INFORMATION

Detailed experimental procedures, analytical and computational methods, tables of calculated shielding factors and sequence probabilities for allowed septet and nonet sequences in copolymer **1**, simulated  $^1\text{H}$  NMR spectra over a range of length scales, atomic coordinates for the energy-minimised model shown in Figure 3a, demonstration of the quaternary expression for the fourth-quarter Cantor set and its generalisation to non-integer bases, full details of the crystallographic analysis and structural data (CIF) for complex **3**, CCDC 1503298. For ESI and crystallographic data in CIF or other electronic format see DOI:xxxxx/yyyyy.

## CONFLICTS OF INTEREST

There are no conflicts of interest to declare.

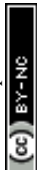
## ACKNOWLEDGEMENTS

We thank Professor Kenneth Falconer of the University of St Andrews and Professor Ricardo Grau Abalo of the Universidad Central de Las Villas for valuable discussions. The work was supported by the Engineering and Physical Research Council of the UK (grant numbers EP/E00413X/1 and EP/G026203/1) and by the European Union (Marie Skłodowska-Curie Network *EURO-SEQUENCES*, grant number 642083). We thank Diamond Light Source Ltd, AWE plc, Cytec Aerospace Materials UK Ltd, and the University of Reading for PhD studentship funding.



## REFERENCES

1. H. F. Judson, *The Eighth Day of Creation*. Cape, London, **1979**.
2. F. H. C. Crick, *Symp. Soc. Exptl. Biol.*, 1958, **12**, 138–163.
3. M. Nirenberg, *Trends Biochem. Sci.*, 2004, **29**, 46–54.
4. D. Thieffry and S. Sarkar, *Trends Biochem. Sci.*, 1998, **23**, 312–316.
5. C. R. Dawkins, *The Blind Watchmaker*, Longmans, London, **1986**, p.115.
6. H. M. Colquhoun and J.-F. Lutz, *Nat. Chem.*, 2014, **6**, 455–456.
7. T. T. Trinh, C. Laure and J.-F. Lutz, *Macromol. Chem. Phys.*, 2015, **216**, 1498–1506.
8. N. Zydziak, W. Konrad, F. Feist, S. Afonin, S. Weidner and C. Barner-Kowollik, *Nat. Commun.*, 2016, **7**, 13672.
9. J. C. Barnes, D. J. C. Ehrlich, A. X. Gao, F. A. Leibfarth, Y. Jiang, E. Zhou, T. F. Jamison and J. A. Johnson, *Nat. Chem.*, 2015, **7**, 810–815.
10. G. Gody, P. B. Zetterlund, S. Perrier and S. Harrison, *Nat. Commun.*, 2016, **7**, 10514.
11. A. E. Stross, G. Iadevaia, D. Nunez-Villanueva and C. A. Hunter, *J. Am. Chem. Soc.*, 2017, **139**, 12655–12663.
12. G. De Bo, M. A. Y. Gall, M. O. Kitching, S. Kuschel, D. A. Leigh, D. J. Tetlow and J. W. Ward, *J. Am. Chem. Soc.*, 2017, **139**, 10875–10879.
13. G. J. Chaitin, *J. ACM*, 1974, **21**, 403–421.
14. J. E. R. Barrios, *Principia*, 2015, **19**, 121–146.





15. B. B. Mandelbrot, *Fractals: Form, Chance and Dimension*. W. H. Freeman and Co., San Francisco, **1977**.
16. S. C. Zimmerman, M. Mrksich and M. Baloga, *J. Am. Chem. Soc.*, 1989, **111**, 8528–8530.
17. H. M. Colquhoun and Z. Zhu, *Angew. Chem., Int. Ed.*, 2004, **43**, 5040–5045.
18. H. M. Colquhoun, Z. Zhu, C. J. Cardin, Y. Gan and M. G. B. Drew, *J. Am. Chem. Soc.*, 2007, **129**, 16163–16174.
19. Z. Zhu, C. J. Cardin, Y. Gan, and H. M. Colquhoun, *Nat. Chem.*, 2010, **2**, 653–660.
20. Z. Zhu, C. J. Cardin, Y. Gan, C. A. Murray, A. J. P. White, D. J. Williams and H. M. Colquhoun, *J. Am. Chem. Soc.*, 2011, **133**, 19442–19447.
21. P. Lazzeretti, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2000, **36**, 1–88.
22. S. Klod and E. Kleinpeter, *J. Chem. Soc., Perkin Trans. 2*, **2001**, 1893–1898.
23. J. Hutchinson, *Indiana Univ. J. Math.*, 1981, **30**, 713–747.
24. J. K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, Wiley, New York, **1990**.
25. H-O. Peitgen, H. Jurgens and D. Saupe, *Chaos and Fractals*. Springer-Verlag, Berlin, **1992**.
26. K. J. Falconer, *Fractals: A Very Short Introduction*. O.U.P., Oxford, **2013**, p.116.
27. G. R. Newkome, P. Wang, C. N. Moorefield, T. J. Cho, P. P. Mohapatra, S. Li, S-H. Hwang, O. Lukoyanova, L. Echegoyen, J. A. Palagallo, V. Iancu and S-W. Hla, *Science* 2006, **312**, 1782–1785.



28. M. Wang, C. Wang, X-Q. Hao, J. Liu, X. Li, C. Xu, A. Lopez, L. Sun, M. P. Song, H. B. Yang and X. Li, *J. Am. Chem. Soc.*, 2014, **136**, 6664–6671.
29. R. Sarkar, K. Guo, C. N. Moorefield, M. J. Saunders, C. Wesdemiotis and G. R. Newkome, *Angew. Chem, Int. Ed.*, 2014, **53**, 12182–12185.
30. J. Shang, Y. Wang, M. Chen, J. Dai, X. Zhou, J. Kuttner, G. Hilt, X. Shao, M. Gottfried and K. Wu, *Nat. Chem.*, 2015, **7**, 389–393.
31. M. P. Parker, C. A. Murray, L. R. Hart, B. W. Greenland, W. Hayes, C. J. Cardin and H. M. Colquhoun, *Cryst. Growth Des.*, 2018, **18**, 386–392.
32. H. J. S. Smith, *Proc. London Math. Soc.*, 1875, **6**, 140–153.
33. G. Cantor, *Math. Ann.*, 1883, **21**, 545–591.
34. G. Cantor, *Acta Math.*, 1884, **4**, 381–392.
35. J. F. Fléron, *Math. Mag.*, 1994, **67**, 136–140.
36. R. C. Archibald, *Amer. Math. Monthly*, 1918, **25**, 189–193.
37. H. M. Colquhoun, D. J. Williams and Z. Zhu, *J. Am. Chem. Soc.*, 2002, **134**, 13346–13347.
38. H. M. Colquhoun, Z. Zhu, D. J. Williams, M. G. B. Drew, C. J. Cardin, Y. Gan, A. G. Crawford and T. B. Marder, *Chem. Eur. J.*, 2010, **16**, 907–918.

